Time- and ensemble-averaged direct NOE restraints

Alexandre M.J.J. Bonvin, Rolf Boelens and Robert Kaptein*

Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

Received 18 August 1993
Accepted 22 October 1993

Keywords: Direct NOE refinement; Time averaging; Ensemble averaging; DINOSAUR

SUMMARY

NMR data are collected as time- and ensemble-averaged quantities. Yet, in commonly used methods for structure determination of biomolecules, structures are required to satisfy simultaneously a large number of constraints. Recently, however, methods have been developed that allow a better fit of the experimental data by the use of time- or ensemble-averaged restraints. Thus far, these methods have been applied to structure refinement using distance and J-coupling restraints. In this paper, time and ensemble averaging is extended to the direct refinement with experimental NOE data. The implementation of time- and ensemble-averaged NOE restraints in DINOSAUR is described and illustrated with experimental NMR data for crambin, a 46-residue protein. Structure refinement with both time- and ensemble-averaged NOE restraints results in lower R-factors, indicating a better fit of the experimental NOE data.

Structure determination by NMR is based on the collection of a large number of constraints, typically obtained from NOE and J-coupling data. These constraints are used to generate and refine a set of structures. Although NMR experimental data are collected as time- and ensemble-averaged quantities, each structure is required to satisfy simultaneously all the constraints. However, generally structures fail to satisfy all the NMR constraints at the same time and may be forced into unrealistic conformations. Solutions to this problem have been proposed by the introduction of time- (Torda et al., 1989, 1990) and ensemble-averaged NMR constraints (Scheek et al., 1991; Kemmink et al., 1993). Time averaging and its implementation in restrained molecular dynamics (MD) simulations have been described and applied first for NOE-derived distances (Torda et al., 1989; Pearlman and Kollman, 1991; Schmitz et al., 1992) and more recently for J-coupling data (Torda et al., 1993). Here we extend the application of time- and ensemble-averaged restraints to the direct refinement against experimental NOE data using relaxation matrix methods.

The experimental NOE intensities can be introduced directly as constraints in a refinement procedure (Yip and Case, 1989), using a penalty function of the form:

*To whom correspondence should be addressed.
\[ V_{\text{NOE}} = \sum w_{ij} \left[ (f A_{ij}^{\text{theo}})^x - (A_{ij}^{\text{exp}})^x \right]^y \]  

(1)

where \( A_{ij}^{\text{exp}} \) and \( A_{ij}^{\text{theo}} \) represent the experimental and theoretical NOE intensities between protons i and j, respectively, \( f \) is a scaling factor and \( w_{ij} \) a weighting function. Often a quadratic potential is used \( (y = 2) \), directly with the NOE intensities \( (x = 1) \) (Yip and Case, 1989; Baleja et al., 1990; Bonvin et al., 1991; Mertz et al., 1991) or with their sixth-root \( (x = 1/6) \) (Nilges et al., 1991) or inverse sixth-root \( (x = -1/6) \) (Stawarz et al., 1992), the last two definitions being closer to a standard distance-based potential. When averaging is introduced, the theoretical intensities in Eq. 1 should be replaced by their time or ensemble averages, denoted as \( \overline{A}_{ij}^{\text{theo}} \). Instead of intensities calculated from a single static structure, now the averaged intensities are required to satisfy the experimental constraints.

In the case of time averaging, for which the time course of an MD simulation can be used, the averaged NOEs, \( \overline{A}_{ij}^{\text{theo}}(t) \) are given by:

\[
\overline{A}_{ij}^{\text{theo}}(t) = \frac{1}{t} \int_0^t A_{ij}^{\text{theo}}(t') \, dt'
\]

(2)

Equation 2 corresponds to the true average, which, as the time increases, might become less sensitive to instantaneous fluctuations in the MD simulation (Torda et al., 1989). To avoid this problem, Torda et al. (1989) introduced an exponentially decaying memory function with time constant \( \tau \) in the summation over the time in Eq. 2, which becomes:

\[
\overline{A}_{ij}^{\text{theo}}(t) = \left[ \tau(1 - \exp(-t/\tau)) \right]^{-1} \int_0^t \exp(-t'/\tau) \, A_{ij}^{\text{theo}}(t - t') \, dt'
\]

(3)

If the simulation time \( t \) is much longer than the time constant of the exponential \( \tau \), a practical way to calculate the average of Eq. 3, suitable for implementation in MD algorithms, is given by (Torda et al., 1989):

\[
\overline{A}_{ij}^{\text{theo}}(t) = \left[ 1 - \exp(-\Delta t/\tau) \right] A_{ij}^{\text{theo}}(t) + \exp(-\Delta t/\tau) \, A_{ij}^{\text{theo}}(t - \Delta t)
\]

(4)

where \( \Delta t \) is the time step of the integrator in the MD simulation. The true average defined in Eq. 2 can, however, be used for the analysis of the MD trajectories.

For ensemble-averaged NOE restraints, the theoretical intensities are given by

\[
\overline{A}_{ij}^{\text{theo}} = \sum_{k=1}^{n_{\text{conf}}} p_k A_{ij}^{\text{theo}}(k) \quad \text{with} \quad \sum_{k=1}^{n_{\text{conf}}} p_k = 1
\]

(5)

where \( p_k \) gives the probability of the conformer \( k \). Ideally, a Boltzmann weighting should be chosen for the probabilities. It is, however, not possible to obtain the free energy to calculate the Boltzmann factor in the course of a simulation; thus, some other weighting function has to be used. Equation 5 requires the computation of the theoretical NOE intensities for all conformers in the ensemble, which is one of the time-consuming steps of direct NOE refinement. However, if